



King's Research Portal

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Ellis, S., & Reader, A. J. (2018). Penalised Maximum Likelihood Simultaneous Longitudinal PET Image Reconstruction with Difference Image Priors. *Medical Physics*.

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Penalised Maximum Likelihood Simultaneous Longitudinal PET Image Reconstruction with Difference-Image Priors

Sam Ellis^{1, a)} and Andrew J. Reader¹

School of Biomedical Engineering and Imaging Sciences, King's College London, King's Health Partners, St Thomas' Hospital, London, SE1 7EH

(Dated: April 19, 2018)

Purpose: Many clinical contexts require the acquisition of multiple positron emission tomography (PET) scans of a single subject, for example to observe and quantify changes in functional behaviour in tumours after treatment in oncology. Typically, the datasets from each of these scans are reconstructed individually, without exploiting the similarities between them. We have recently shown that sharing information between longitudinal PET datasets by penalising voxel-wise differences during image reconstruction can improve reconstructed images by reducing background noise and increasing the contrast-to-noise ratio of high activity lesions. Here we present two additional novel longitudinal difference-image priors and evaluate their performance using 2D simulation studies and a 3D real dataset case study.

Methods: We have previously proposed a simultaneous difference-image-based penalised maximum likelihood (PML) longitudinal image reconstruction method that encourages sparse difference images (DS-PML), and in this work we propose two further novel prior terms. The priors are designed to encourage longitudinal images with corresponding differences which have i) low entropy (DE-PML), and ii) high sparsity in their spatial gradients (DTV-PML). These two new priors and the originally proposed longitudinal prior were applied to 2D simulated treatment response [¹⁸F]fluorodeoxyglucose (FDG) brain tumour datasets and compared to standard maximum likelihood expectation-maximisation (MLEM) reconstructions. These 2D simulation studies explored the effects of penalty strengths, tumour behaviour, and inter-scan coupling on reconstructed images. Finally, a real two-scan longitudinal data series acquired from a head and neck cancer patient was reconstructed with the proposed methods and the results compared to standard reconstruction methods.

Results: Using any of the three priors with an appropriate penalty strength produced images with noise levels equivalent to those seen when using standard reconstructions with increased counts levels. In tumour regions each method produces subtly different results in terms of preservation of tumour quantification and reconstruction root mean-squared error (RMSE). In particular, in the two-scan simulations, the DE-PML method produced tumour means in close agreement with MLEM reconstructions, while the DTV-PML method produced the lowest errors due to noise reduction within the tumour. Across a range of tumour responses and different numbers of scans, similar results were observed, with DTV-PML producing the lowest errors of the three priors and DE-PML producing the lowest bias. Similar improvements were observed in the reconstructions of the real longitudinal datasets, although imperfect alignment of the two PET images resulted in additional changes in the difference image that affected the performance of the proposed methods.

Conclusion: Reconstruction of longitudinal datasets by penalising difference images between pairs of scans from a data series allows for noise reduction in all reconstructed images. An appropriate choice of penalty term and penalty strength allows for this noise reduction to be achieved while maintaining reconstruction performance in regions of change, either in terms of quantification of mean intensity via DE-PML, or in terms of tumour RMSE via DTV-PML. Overall, improving the image quality of longitudinal datasets via simultaneous reconstruction has the potential to improve upon currently used methods, allow dose reduction, or reduce scan time while maintaining image quality at current levels.

Keywords: positron emission tomography; image reconstruction; longitudinal reconstruction; treatment response; longitudinal studies

I. INTRODUCTION

Positron emission tomography (PET) is widely used to observe and measure functional and metabolic behaviour *in vivo*. In a number of contexts, multiple PET

scans are performed in order to observe and/or measure changes in these biological processes over long time scales, ranging from days up to several months. One of the most important examples of such longitudinal PET protocols is in cancer imaging, where the response of a tumour to a treatment is often of interest¹. Within this area, multiple PET scans are used for a wide range of purposes such as predicting treatment response², guiding future treatment³, or assessing the efficacy of new

^{a)}Corresponding author: sam.ellis@kcl.ac.uk

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:

10.1002/mp.12937

This article is protected by copyright. All rights reserved.

biomarkers in reflecting tumour response⁴. Such studies have been performed for various cancer treatments, such as chemotherapy⁵, radiotherapy⁶, and immunotherapy⁷. The number of scans that are performed in longitudinal oncological protocols varies greatly; although two to five scans are typical^{2,6,8}, more than ten can be acquired in specific contexts⁷. Analysis of longitudinal changes throughout these image series is also varied, ranging from qualitative methods², through semi-quantitative measures like standardised uptake values (SUVs)^{3,5,6}, to full quantification of tumour metabolism, e.g. by Patlak analysis^{8,9}. Reported SUVs include the maximum SUV measured within the tumour⁵, the mean value within the tumour⁶, and in some cases the SUV is compared to a background region such as the liver for increased stability^{4,7}.

For the majority of these longitudinal PET studies, images are typically reconstructed with iterative reconstruction methods such as maximum likelihood expectation-maximisation (MLEM¹⁰) or, more commonly, an accelerated version of MLEM known as ordered subsets expectation-maximisation, or OSEM¹¹. Iterative methods can produce PET images that are superior in quality to analytic methods such as filtered back-projection (FBP) by incorporating more realistic models of the acquisition process and by taking into account the Poisson distributed nature of the acquired counts¹².

These iterative methods attempt to find the set of parameters (most commonly voxel intensities) which produce expected data that is closest to the measured data, given some model of the acquisition process. However, due to the counts-limited nature of PET data, they produce noisy images at convergence. In clinical practice, this problem is solved by terminating the reconstruction early, before the noise becomes excessive. However, using unconverged image estimates can impact on quantification by introducing bias into intensity values. An alternative approach is to regularise the PET reconstruction to avoid noisy images. One of the most common methods of regularising PET image reconstructions is to use a Bayesian or penalised maximum likelihood framework. These methods allow the introduction of prior expectations into the reconstruction, and allow a trade-off between how well the reconstructed image matches the measured data (data consistency), and how well it matches the prior expectations. Since noise manifests itself as large differences between neighbouring voxels, a common technique for reducing PET image noise via regularisation methods is to include a penalisation of neighbourhood differences in the prior expectations. However, large neighbourhood differences are to be expected over valid image edges and the penalisation is not desired in these regions of the image. It is possible to define isotropic, edge-preserving reconstruction priors that use total-variational methods^{13–15}, but these methods can produce artificial-looking, piecewise constant images that may not reflect the true nature of the activity distribution.

Another option for edge-preservation in regularised PET image reconstruction is to use a prior source of knowledge to define which voxels are expected to appear similar, thereby switching off the PET regularisation over valid edges. A promising source of such information is from magnetic resonance (MR) images of the same patient, since MR images often have superior noise and resolution properties to PET. There have been many studies into the use of MR-based priors in PET reconstruction^{16–22}, but one key remaining question is in ensuring that MR-specific features are not imprinted upon the PET image.

The advent of simultaneous PET-MR systems provides convenient MR images for guiding PET image reconstruction; due to the simultaneity of the acquisitions, the inter-modality registration problem is made simpler. Furthermore, simultaneous PET-MR image reconstruction has been proposed to allow both modalities to influence the reconstruction of the other, allowing improved PET images along with faster MR acquisition times^{23–25}. However, again, there is a danger of unwanted cross-talk between the imaging modalities degrading the reconstructed images.

Inspired by joint simultaneous PET-MR reconstruction methods, we aim to couple longitudinal PET datasets during the reconstruction process to improve the images from all scans. Our proposed methods operate in the difference image domain in order to produce PET images that are natural looking and of a superior quality. We previously proposed the simultaneous reconstruction of longitudinal PET datasets²⁶, where noise reduction was achieved by penalising voxel differences between longitudinal PET scans. Our results showed that this method allows for the reduction of image-wide reconstruction errors and noise levels while maintaining the appearance of regions of change, resulting in higher contrast-to-noise ratios for lesions in inserted lesion brain scans. However, the method was restricted to two scans with the same background intensity levels. In this work we aim to extend the simultaneous longitudinal reconstruction framework to a general number of scans with arbitrary activity levels. Additionally, we present two new longitudinal priors that have been designed to improve reconstruction performance within lesions. These new simultaneous longitudinal reconstruction methods are characterised using 2D simulation studies before being applied to real datasets, where two PET images were acquired in the context of a head and neck treatment response study.

II. MATERIALS AND METHODS

A. Theory

Let θ be a vector containing the intensities of a PET image so that θ_j is the model of the mean of the number of emissions (proportional to activity concentration) from

the j th voxel. In this work, we consider a set of longitudinal PET images, denoted as $\{\boldsymbol{\theta}_s\}$, where $s = 1, \dots, S$ indicates the scan number. In this notation, the intensity of the j th voxel of the s th longitudinal scan is denoted $\theta_{j,s}$.

The joint reconstruction of multiple PET datasets is achieved by formulating the following penalised maximum likelihood (PML) objective function:

$$\Phi(\{\boldsymbol{\theta}_s\}) = \sum_{s=1}^S L(\boldsymbol{\theta}_s | \mathbf{y}_s) - \beta U(\{\boldsymbol{\theta}_s\}) \quad (1)$$

where \mathbf{y}_s is the measured data vector (for example, the number of recorded counts in each sinogram bin) for scan s , $L(\boldsymbol{\theta}_s | \mathbf{y}_s)$ is the Poisson log-likelihood of image $\boldsymbol{\theta}_s$ producing \mathbf{y}_s , $U(\{\boldsymbol{\theta}_s\})$ is a penalty function in terms of all the longitudinal images, and β is a hyperparameter that controls the strength of the regularisation.

In this work, we apply penalties to the set of difference images, $\{\tilde{\boldsymbol{\delta}}_{sk}\}$, given by

$$\tilde{\boldsymbol{\delta}}_{sk} = \tilde{\boldsymbol{\theta}}_k - \tilde{\boldsymbol{\theta}}_s \quad (2)$$

where $\tilde{\boldsymbol{\theta}}_s = \alpha_s \boldsymbol{\theta}_s$ represents the image $\boldsymbol{\theta}_s$ normalised so that voxels within regions which have not changed metabolically between scans yield differences of zero. In general, the penalty term is written as:

$$U(\{\boldsymbol{\theta}_s\}) = \sum_{s=1}^S \sum_{k=1}^S w_{sk} u(\tilde{\boldsymbol{\delta}}_{sk}) \quad (3)$$

Explicitly, given S scans, the S^2 difference images can be calculated and the total penalty is the weighted sum of the function u applied to all possible difference images. The weights w_{sk} control the coupling between scans in the reconstruction; setting $w_{sk} = 0$ ensures that there is no similarity encouraged directly between scans s and k .

In this work, in order to seek the maximum of Equation (1), we employ the one-step-late iterative reconstruction method proposed by Green²⁷. This results in the following update equation:

$$\theta_{j,s}^{(\nu+1)} = \frac{\theta_{j,s}^{(\nu)}}{\sum_{i=1}^I P_{ij,s} + \beta \left. \frac{\partial U}{\partial \theta_{j,s}} \right|_{\{\boldsymbol{\theta}_s^{(\nu)}\}}} \sum_{i=1}^I \frac{P_{ij,s} y_{i,s}}{\bar{y}_{i,s}(\boldsymbol{\theta}_s^{(\nu)})} \quad (4)$$

where $P_{ij,s}$ are the elements of the system matrix for scan s , ν is the iteration number, and $\bar{\mathbf{y}}_s(\boldsymbol{\theta}_s^{(\nu)})$ is the expected mean data vector given the image $\boldsymbol{\theta}_s$, according to the affine relationship:

$$\bar{y}_{i,s}(\boldsymbol{\theta}_s) = \sum_{j=1}^J P_{ij,s} \theta_{j,s} + b_{i,s} \quad (5)$$

where $b_{i,s}$ are the expected additive contributions to the measured data through random coincidences and scattered photons. As shown in equation 4, the one-step-late

algorithm uses the gradient of the penalty function evaluated at the current image estimate to allow a closed form update similar to that of standard MLEM. The disadvantage of this approach is that it is not guaranteed to converge to a global maximum of the objective function and can result in unstable reconstructions when β is too large. Nonetheless, the one-step-late update has been widely used in the PET image reconstruction literature and provides valuable insight into the performance of PML reconstruction methods provided β is not too large.

1. Difference image sparsity prior – DS-PML

Assuming that the only a small image region changes from scan to scan, one possible prior is to expect the difference images to be sparse. In the inverse problems and parameter estimation literature, the ℓ_1 -norm is often used as a convex surrogate for the ℓ_0 -norm to encourage sparsity in some domain. Therefore, the difference image sparsity (DS) prior for longitudinal PET image reconstruction uses the following form for $u(\tilde{\boldsymbol{\delta}}_{sk})$:

$$u_{\text{DS}}(\tilde{\boldsymbol{\delta}}_{sk}) = \sum_{j=1}^J \sqrt{\tilde{\delta}_{j,sk}^2 + \varepsilon^2} \quad (6)$$

where ε is a small value introduced to ensure that the prior is differentiable at all values. In previous experiments²⁶ a similar longitudinal prior was shown to reduce PET reconstruction errors by lowering image noise in regions of no change, while producing small biases in longitudinally changing tumours.

2. Difference image entropy prior – DE-PML

For longitudinal PET images consisting of a number of distinct tissue classes (e.g. white matter, grey matter, tumour), it is expected that each one of these tissues varies in activity across the longitudinal scans in a consistent way, e.g. that all grey matter voxels change in intensity by the same amount. This can be formulated as assuming that the difference image should have a low entropy. The use of information-based priors has previously been explored in the field of MR-guided PET image reconstruction¹⁶, but to our knowledge, they have not been incorporated into longitudinal PET reconstructions or applied to difference images.

Given a set of parameters $\{\theta_j\}$, the probability density function (PDF) of a continuous random variable, X , can be estimated as $\hat{p}(x; \boldsymbol{\theta})$ using the Parzen window method, or kernel density estimation:

$$\hat{p}(x; \boldsymbol{\theta}) = \frac{1}{J} \sum_{j=1}^J \phi(x - \theta_j). \quad (7)$$

With this formula we can estimate the entropy of an image vector according to:

$$\hat{H}(X; \theta) = - \int \hat{p}(x; \theta) \log \hat{p}(x; \theta) dx. \quad (8)$$

As previously mentioned, we want to encourage solutions which have a low entropy difference image, i.e:

$$u_{\text{DE}}(\tilde{\delta}_{sk}) = \hat{H}(X, \tilde{\delta}_{sk}). \quad (9)$$

Note that in practice we replace the continuous variable X with a discrete approximation, and the integral becomes a sum. Using this formulation, the gradient of the entropy prior can be calculated¹⁶ and used for the one-step-late reconstruction.

3. Difference image total variation prior – DTV-PML

The total variation (TV) of a 2D image can be defined as²⁸

$$\text{TV}(\theta) = \sum_{x,y} \sqrt{(\theta_{x+1,y} - \theta_{x,y})^2 + (\theta_{x,y+1} - \theta_{x,y})^2 + \varepsilon^2} \quad (10)$$

where $\theta_{x,y}$ here denotes the image indexed in each spatial dimension separately. Here we propose that we expect the difference images to have sparse spatial gradient images, i.e. that the TV of the difference images is low. Therefore, the difference image TV prior (DTV) uses the following for $u(\tilde{\delta}_{sk})$:

$$u_{\text{DTV}}(\tilde{\delta}_{sk}) = \text{TV}(\tilde{\delta}_{sk}). \quad (11)$$

The TV penalty has previously been used in single-dataset PET reconstruction²⁸, but to the best of the authors' knowledge, it has never been applied to difference images between longitudinal PET scans.

B. 2D ground truth images and data simulation

Realistic brain phantom software²⁹ was used to create 2D [¹⁸F]fluorodeoxyglucose (FDG) ground truth images comprising four segmented intensity regions: background, white matter, grey matter, and tumour. Background voxels were assigned activities of 0 and the ratio between grey and white matter intensities was 4:1. The tumour was a variable uniform circular region on the boundary between the left middle frontal gyrus and white matter, characterised by its radius and intensity. Ground truth images had an array size of 512×512 with a pixel size of $0.5 \times 0.5 \text{ mm}^2$.

To transform between image and data spaces, the Radon and adjoint Radon transforms were used for projection and backprojection respectively. In order to generate the datasets for reconstruction, the ground truth

images were blurred by a 4.3 mm Gaussian kernel before forward projecting. Sinogram space attenuation factors were calculated from an attenuation map comprising background ($\mu = 0 \text{ mm}^{-1}$), soft tissue ($\mu = 0.0096 \text{ mm}^{-1}$), and bone ($\mu = 0.0172 \text{ mm}^{-1}$)³⁰. Random events were approximated by a uniform sinogram, and scattered coincidences were approximated by smoothing the projected ground truth in each case. The scatters and randoms were scaled and added to the data to produce noise-free sinograms which contained 20% scatter and 20% randoms. Noise-free and non-resolution degraded data were reconstructed with filtered backprojection (FBP) to provide reference images in the same space as the noisy reconstructions.

C. Experiment 1: Hyperparameter selection

Two ground truth datasets were generated as outlined above, one baseline (PET1) and one follow-up (PET2). The PET1 tumour had a radius of 15 mm and a tumour to white matter intensity ratio of 8:1, and the PET2 tumour was 8 mm with a corresponding ratio of 6:1. The two images were scaled in order to set the total expected number of counts in the PET1 dataset to approximately 4 million. In addition, a second, double-counts pair of noise-free datasets were generated with double the intensity to provide a low-noise benchmark.

100 noisy realisations of each of the datasets (PET1, PET2, double-counts PET1, and double-counts PET2) were then produced by introducing Poisson noise into the noise-free datasets. These datasets were reconstructed with a variety of methods, all run for 200 iterations, and all using the original attenuation, scatters and randoms sinograms from the simulation process. The datasets with an average of approximately 4 million counts per sinogram were reconstructed with the following methods: 1. independent dataset MLEM, 2. DS-PML, 3. DE-PML, and 4. DTV-PML. The noisy double-counts PET1 and PET2 were reconstructed with independent dataset MLEM to demonstrate the image quality achievable when a higher level of counts are recorded.

For each of the longitudinal difference-image prior reconstruction methods, the scan-to-scan weighting matrix w_{sk} was set to a value of 1 for all s, k , as were all normalising factors α_s . The prior specific parameters were as follows:

a. *DS-PML*: β values were varied between 7.5×10^{-5} to 15×10^{-4} and ε was set to 1×10^{-6} .

b. *DE-PML*: A Gaussian window of standard deviation equal to 0.35 (arb. intensity units) was used to estimate the PDF of the difference images. During each iteration of the DE-PML algorithm, a 100-level discretisation was used for approximating the continuous PDF. In addition, a mask was used to ensure that entropy was only calculated in the brain itself since the background naturally has low entropy. β values were varied between

0.15 and 3.

c. DTV-PML: β values were varied between 4×10^{-5} to 8×10^{-4} and ε was set to 1×10^{-6} .

In addition, a longitudinal smooth was applied to the MLEM reconstructions, using a two-point Gaussian kernel parameterised by the standard deviation, σ , so that with $\sigma = 0$ the images are unchanged, and with $\sigma = +\infty$ the outputs are just the average of the input images.

To evaluate the quality of the reconstructed images, the 100 noisy realisations of the datasets were reconstructed as described above and the following regional metrics were calculated.

For a set of images, $\{\theta_{jn}\}$, with n indexing noisy realisations, the percentage root square mean error relative to a reference image $\{\theta_j^{\text{Ref}}\}$ for a given region Ω is calculated as:

$$\% \text{RMSE} = \frac{100\%}{N_\Omega} \sum_{j \in \Omega} \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (\theta_{jn} - \theta_j^{\text{Ref}})^2}}{\theta_j^{\text{Ref}}} \quad (12)$$

where N_Ω is the number of voxels in Ω and $N (= 100)$ is the number of noisy realisations. %RMSE values were calculated for PET1 and PET2 images using the noise-free FBP images as the corresponding reference images. Ω was selected as a 226-voxel circular region of interest (ROI) encompassing the 15 mm tumour and some surrounding tissue. Note that the same region was used in both PET1 and PET2 analyses so that the PET2 image, with the smaller tumour, contained more background within Ω .

The mean intensities within the PET1 and PET2 tumours were also calculated for each reconstruction, and the average of these across all noisy realisations calculated to provide a measure of the effects of the regularisation penalties on tumour quantification.

To measure the background noise present in the images, the coefficient of variation (CV, equal to the standard deviation divided by the mean) was calculated in an eroded white matter mask.

Finally, bias and standard deviation maps were produced by calculating the bias (relative to the noise-free FBP images) and standard deviation of each voxel across the $N = 100$ noisy realisations.

From the results of Experiment 1, hyperparameters for each of the reconstruction methods were selected and used for the rest of the 2D simulation experiments, where imaging conditions and counts levels per scan were nominally the same as in Experiment 1.

D. Experiment 2: Varying tumour response

Experiment 1 considered only one example of a tumour response. To evaluate the performance of the presented difference-image prior methods for a range of pre- and post- treatment behaviours, we performed a second experiment in which tumour radius and intensity were varied and the performance of the reconstruction methods

with fixed β was assessed. Initially keeping the tumour intensities the same as listed in Section II C, the PET1 and PET2 tumour radii were varied between 3 mm and 15 mm in steps of 2 mm. 100 noisy realisations of simulated PET data were generated for each radius pair using the same methodology and counts levels as in Experiment 1, and reconstructed with 200 iterations of each method with the selected hyperparameters. For all PET1 and PET2 reconstructions for all tumour pairs, %RMSE was calculated in the same 226-voxel circular ROI used in Experiment 1.

Following this, the radii of the two tumours were kept the same as listed in Section II C, with the intensities varied between 2 and 8 (relative to a nominal white matter intensity of 1). Using 200 iterations of each reconstruction method, with 100 noisy realisations per dataset again, %RMSE was calculated in the tumour ROI.

E. Experiment 3: Five-scan simulation study

Experiments 1 and 2 used the simplifying case of a two-scan protocol. To investigate the effect of the inter-scan weights w_{sk} , a five-scan case was simulated. Five ground truth images and datasets were created in accordance with Section II B. The tumour radii were: 15 mm, 8 mm, 4 mm, 3 mm, and 5 mm, and the respective intensities were 8, 6, 4, 3, and 6 (relative to white matter). The counts level of each scan was the same as that used in Experiments 1 and 2, i.e. approximately 4 million counts per scan, and 100 noisy realisations of each dataset were generated.

The five datasets were reconstructed with each of the difference-image prior methods. The w_{sk} array is now rewritten so that:

$$w_{sk} = \kappa_s \exp \left(-\frac{(\Delta_{sk})^2}{2\sigma_w^2} \right) \quad (13)$$

where σ_w is a parameter that can be varied from 0 (i.e. $w_{sk} = 1$ iff $s = k$) to $+\infty$ (i.e. $w_{sk} = w \forall s, k$), and $\Delta_{sk} = \min(|s - k|, |s - k - S|, |s - k + S|)$ is the cyclic longitudinal distance between scans s and k . Using a cyclic distance results in all scans having the same total weighting with other scans, ensuring that, for example, scans 1 and 5 are not regularised less just by virtue of having no preceding or succeeding scan. κ_s is a normalising factor used to control the sum of the rows of the weights array. In this experiment we choose κ_s so that $\sum_k w_{sk} = 2$, in agreement with the total weight used in the previous experiments where w_{sk} was $\mathbf{1}_{2 \times 2}$.

The σ_w parameter was varied between 0.4 and 8, with a final σ_w of ∞ also included to provide a case comparable to those used in Experiments 1 and 2. β values for the difference-image priors were set to the same values as used in Experiment 2, based on the results of Experiment 1.

In order to provide comparable images at different image noise levels, the five noise-free longitudinal datasets

were scaled to different counts levels, used to generate 100 noisy, high-counts datasets, and then reconstructed with 200 iterations of MLEM. These counts levels ranged from $1\times$ to $5\times$ the original level, to correspond to the maximum noise reduction expected from the DS-PML, DE-PML and DTV-PML methods.

Reconstructed images for all methods were analysed by calculating mean tumour intensities, white matter CV values, and mean squared error (MSE, equal to the sum of the squared bias and squared standard deviation) maps.

F. Experiment 4: Application to real data for head and neck cancer

In the final experiment, the proposed priors were applied to a real longitudinal pair of [^{18}F]FDG datasets acquired from a head and neck cancer patient in a Siemens mCT time-of-flight (TOF) PET-CT scanner (Siemens Healthcare, Erlangen, Germany). The patient exhibited a highly-uptaking tumour in the right side of the lower jaw, which reduced in size and uptake between the two scans (an interval of 85 days). The injected activity of each of the two scans was 340 MBq and 357 MBq, resulting in total prompt counts of 78.5×10^6 and 76.7×10^6 respectively in sinograms acquired in the head bed position. Note that although the data were acquired in TOF format, sinograms were rebinned to non-TOF for this experiment.

As well as the prompts sinograms, computed tomography (CT) images were acquired for attenuation correction, and delayed sinograms were acquired for randoms estimation. The vendor-supplied e7 tools software was used to calculate attenuation correction factors and the randoms sinograms for each dataset to be used in the reconstructions. Scatter correction was not performed.

To allow a comparison between reconstructions with the proposed methods and double-counts images as performed in Experiments 1 to 3, the counts levels in the original mCT sinograms were reduced by 50% by randomly removing individual counts with a probability of 0.5. The randoms estimations were scaled by a factor of 0.5 accordingly.

The counts-reduced longitudinal datasets were then reconstructed with DS-PML, DE-PML, and DTV-PML, using projectors based on the mCT geometry³¹. To reduce reconstruction time, these reconstruction methods were accelerated with an ordered subsets implementation¹¹. The same counts-reduced datasets were also reconstructed with standard OSEM. All reconstructions were run with 5 iterations of 21 subsets reconstructing into an image grid size of $400 \times 400 \times 109$, with voxel sizes of $2.036 \times 2.036 \times 2.027 \text{ mm}^3$. Each reconstruction used resolution modelling with a Gaussian kernel of full-width-at-half-maximum of 3 mm.

To address the inherent mis-alignment between the datasets due to differing patient position in the two acquisitions, non-rigid registration was performed with the

two CT-based attenuation maps utilising a demons-based registration (MATLAB, MathWorks, MA, USA). Using the baseline scan as the reference image space, the calculation of the penalty gradients for each difference-image prior, as required in equation 4, was performed by applying the non-rigid alignment fields to the current PET2 image before calculation of the penalty gradients. This produced the penalty gradient in the same image space as the PET1 scan, so that the alignment operator estimated from the image registration process was then used to transform the penalty gradients back into the PET2 space. Note that with this methodology each of the PET datasets are reconstructed in their respective native image spaces; the only time alignment is applied is in the calculation of the penalty gradient terms.

For the difference-image prior methods the weights w_{sk} were 1 for all s and k , and α_s were given by T_1/T_s where T_s denotes the total number of prompt counts recorded in scan s . For DS-PML, β was varied between 0 and 0.2 and ε was 1×10^{-6} . For DE-PML, β was varied between 0 and 2000 with a Gaussian Parzen window width of 0.02. Finally, for DTV-PML, β values were varied between 0 and 0.1, and $\varepsilon = 1 \times 10^{-6}$. Whereas in the 2D simulation studies only DE-PML required a masking of voxels for calculation of the penalty gradient, for the real 3D datasets masking was required for all priors due to low sensitivity at the edges of the field-of-view. This penalty mask was calculated using the extent of the patient's head based on the PET1 attenuation map, and cropped to remove the low-sensitivity axial edges of the field-of-view.

The quality of the reconstructed images was assessed by measuring the mean of the tumour and the CV in a background region in both the PET1 and PET2 reconstructed images. To serve as a reference, the original full-counts data were reconstructed into the same image grid, using OSEM with 5 iterations of 21 subsets, and the tumour mean and background CV were measured.

III. RESULTS

A. Experiment 1: Hyperparameter selection

Figure 1 shows the tradeoffs between tumour mean and white matter CV as a function of regularisation hyperparameter for the proposed methods. An ideal method would provide a point that coincides with the double-counts MLEM point with only the same amount of data that is available to the proposed methods. It is clear that compared to a naive longitudinal smooth, the proposed methods are far superior in terms of limiting the bias introduced into the tumour. In particular, the DE-PML method provides high amounts of noise reduction while keeping the tumour mean close to its MLEM value.

The corresponding tradeoff between tumour %RMSE and white matter CV is shown in figure 2. Here the double-counts MLEM reconstructions exhibit both lower

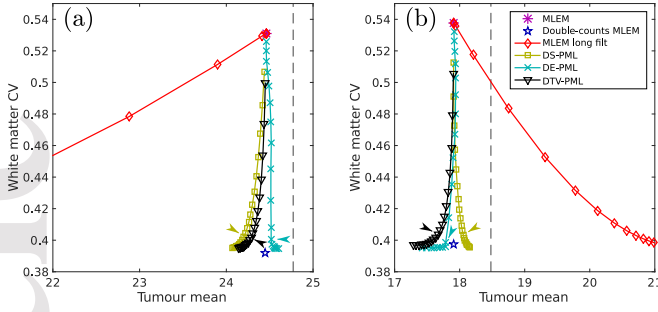


Figure 1. Tumour mean vs white matter CV at 200 iterations as a function of regularisation hyperparameter for (a) PET1 and (b) PET2. For the longitudinally smoothed MLEM reconstructions (red diamonds), the hyperparameter is σ , controlling the width of the filter, and for DS-PML (yellow squares), DE-PML (cyan crosses) and DTV-PML (black triangles), the parameter is β . See Section II C for the ranges of σ and β values used for each method. Also shown are the performances of MLEM (purple asterisks) and MLEM applied to the double-counts data (blue stars). Dashed lines show the ground truth tumour mean values. Note that in both PET1 and PET2 cases, the longitudinal smooth with $\sigma \rightarrow \infty$ is outside the displayed axes. Arrows indicate the selected β values for each method, as listed in table I, which are used for the remainder of the 2D simulation experiments.

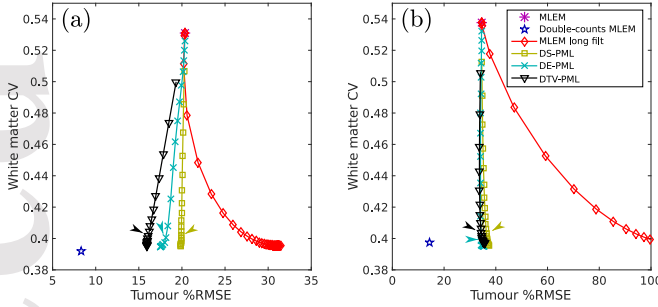


Figure 2. Tumour %RMSE vs white matter CV at 200 iterations as a function of regularisation hyperparameter for (a) PET1 and (b) PET2 reconstructions. Reconstruction methods shown are MLEM, DS-PML, DE-PML, and DTV-PML. Also shown are longitudinally smoothed MLEM (“long fit”) and MLEM reconstructions using double-counts data. Arrows indicate the selected β values listed in table I and used for the remainder of the 2D simulation experiments.

tumour %RMSE and lower white matter CV. The DS-PML method at high β values allows high levels of white matter noise reduction, while approximately keeping tumour %RMSE similar to the MLEM reconstructions. On the other hand, the DE-PML and in particular the DTV-PML methods reduce PET1 tumour error compared to the MLEM reconstructions, at the same time as reducing white matter CV. For the PET2 reconstructions, the three difference-image priors produce similar error levels, due to the region containing a proportionally higher amount of background than in the PET1 case.

To assess the tradeoff between bias and variance that

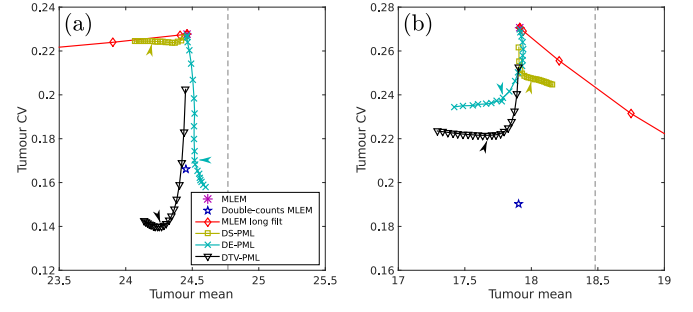


Figure 3. Tumour mean against tumour CV for varying regularisation parameter for (a) PET1 and (b) PET2 reconstructions. DS-PML, DE-PML and DTV-PML methods are shown as a function of β . Also shown are MLEM, MLEM with double-counts data, and longitudinally filtered MLEM, parameterised by the width of the filter, σ . Arrows indicate the β values listed in table I, used in the remainder of the 2D simulation experiments.

gives rise to the %RMSE values in figure 2, the tumour mean was plotted against the tumour CV as a function of β for the presented difference-image prior methods, as shown in figure 3. It was observed that both DTV-PML and DE-PML produced noise reduction in the tumour itself, resulting in the lower %RMSE values observed previously. The DS-PML method performed worse than the other two methods in terms of noise reduction. It is interesting to note that in the PET1 reconstructions, DE-PML and DTV-PML with sufficiently high values of β reduced tumour noise to levels considerably below even the double-counts reconstruction, although for the PET2 reconstructions tumour noise plateaued at a CV level greater than the double-counts case.

Based on figures 1 to 3, the selection of an optimal β value for each method is non-trivial. In practice, there is a tradeoff between all variables shown in those figures, so that any single β value cannot be optimal for all metrics of image quality. Since optimising so many correlated and anti-correlated metrics simultaneously is difficult, and since ultimately any clinical usage of these algorithms would be subject to the requirements of the particular application in question, we did not attempt to select β values based on a quantitative measure. Instead, to demonstrate both the full potential offered by each method and their respective drawbacks, we selected β values that were sufficiently high so as to approach the noise levels of the double-counts MLEM reconstructions, without excessive penalisation. The selected values are indicated by arrows in figures 1 to 3, and listed in table I.

To test the convergence of the presented reconstruction methods, a single noisy realisation of the Experiment 1 data was reconstructed with each of the methods with the β values listed in table I, as well as with MLEM. All reconstructions were run to 1000 iterations, and the value of the corresponding objective function at each iteration number was calculated. Figure 4 shows the normalised

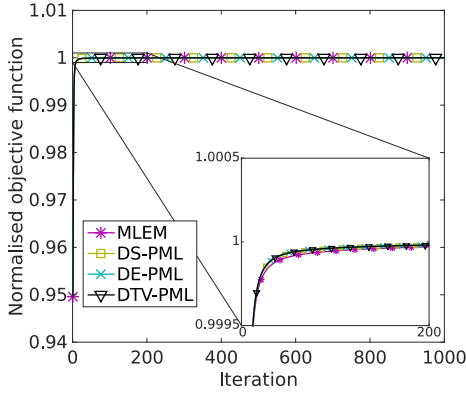


Figure 4. Plot of the objective function as a function of iteration number for a single noisy realisation of the experiment 1 datasets reconstructed with MLEM, DS-PML, DE-PML, and DTV-PML. For each of the PML reconstructions, β was set to the value indicated in table I. Note that for each method, the objective function values have been normalised by the value obtained after 1000 iterations.

objective function as a function of iteration for MLEM and the three PML methods. Observing the progression of the normalised objective function with iteration number, it is apparent that all of the methods converge with a similar trajectory to MLEM, with each one within 99.99% of their 1000-iteration value at 200 iterations.

Figure 5 plots the tumour mean values for the proposed methods with the selected β values alongside the MLEM and double-counts MLEM tumour means. It is clear that introducing a longitudinal prior can bias the estimate of the tumour means, but in this case the effect is small even for the most severe cases (PET2 DTV-PML or both DS-PML reconstructions).

When considering the bias and standard deviation maps (figure 6), it is apparent that all of the proposed methods reduce the voxel-wise variance in background regions due to the noise reduction indicated by figures 1 and 2. However, behaviour in and around the tumour is different for the various methods, with DS-PML maintaining voxel-wise variance at levels similar to MLEM, and with the DE-PML and DTV-PML methods reducing voxel-wise variance in the tumour. In terms of the bias maps, the difference-image prior methods produce similar results to the MLEM reconstructions, but with a slight visible increase in bias in the edges of the PET2 tumour for DE-PML and DTV-PML.

| Algorithm | Selected β value |
|-----------|------------------------|
| DS-PML | 9×10^{-4} |
| DE-PML | 1.8 |
| DTV-PML | 4×10^{-4} |

Table I. β values selected for use in the remainder of the 2D simulation experiments.

B. Experiment 2: Varying tumour response

Figure 7 displays the %RMSE in the tumour region as a function of PET1 and PET2 tumour radius, and figure 8 shows the corresponding results for changing PET1 and PET2 tumour intensities. The CV in background regions is similar to those shown in figures 1 and 2 and so are not replicated.

Firstly, it is apparent that for all tumour sizes doubling the number of counts used in the MLEM reconstruction reduces reconstruction errors, reflecting the results shown in figure 2. In general, a smaller tumour radius and intensity causes higher reconstruction errors in MLEM reconstructions.

When using the difference-image priors, error levels are generally the same as or reduced compared to MLEM reconstructions with the same number of counts when the tumour radii vary (figure 7). In particular, with the difference-image prior methods, the highest errors occur when the change in tumour radius is greatest (c.f. bottom right and top left corners of the heat maps shown in figure 7). For the DS-PML method, these extreme cases have %RMSE levels similar to the MLEM reconstructions, while elsewhere %RMSE is reduced slightly due to noise reduction. For the DE-PML and DTV-PML methods, even at the most extreme radii changes, error levels are lower than the MLEM case, in agreement with figure 1. In some cases these errors approach the double-counts MLEM reconstructions.

For the reconstruction errors in the tumour region as a function of varying tumour intensities (figure 8), similar results are observed, with the DS-PML error levels similar to the MLEM errors, and the DE-PML and DTV-PML errors falling between the MLEM and the double-counts MLEM errors.

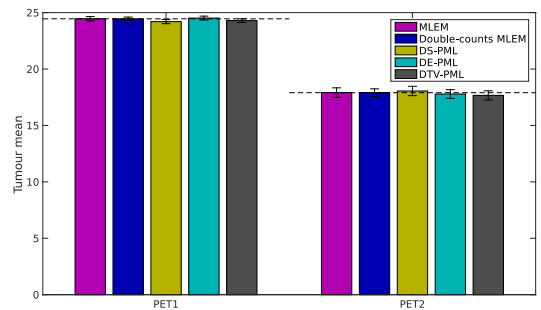


Figure 5. Tumour mean values at 200 iterations for MLEM, double-counts MLEM, DS-PML, DE-PML and DTV-PML. β values for the PML methods are those listed in table I. Error bars show the standard deviation of the measured mean values over the 100 noisy realisations of the data. Dashed lines replicate the double-counts MLEM tumour means for comparison.

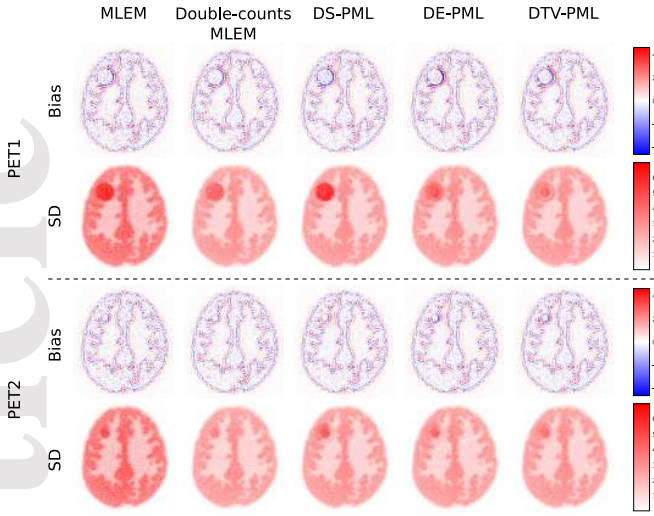


Figure 6. Bias and standard deviation (SD) images calculated over 100 noisy realisations for the PET1 and PET2 reconstructed with MLEM, double-counts MLEM, DS-PML, DE-PML and DTV-PML. β values for the PML methods are the same as those listed in table I.

C. Experiment 3: Five-scan simulation study

Figure 9 shows the tumour mean values and the white matter CV as a function of σ_w for DS-PML, DE-PML and DTV-PML with β values unchanged from Experiment 2. Also shown are the corresponding values for MLEM with a number of recorded counts varied between $1\times$ and $5\times$ the number used for the difference-image prior methods. As σ_w tends towards infinity, the noise is increasingly reduced until at $\sigma_w = \infty$, the noise levels reach a minimum which is determined by the selected β value. Out of the three priors, the DE-PML method gives the least biased tumour mean values, with $\sigma_w = \infty$ providing values comparable to MLEM reconstructions.

Observation of the PET1 and PET2 tumour mean/CV tradeoffs (figure 10) reflects the results seen in figure 3, despite the change of regularisation parameter from β to σ_w . In figure 10, increasing σ_w reduces tumour CV for the DTV-PML method, and to a lesser extent the DE-PML method. DS-PML has a weaker effect on tumour CV; in the PET1 reconstructions, the tumour noise was maintained at levels similar to the $1\times$ counts MLEM reconstructions, despite the introduction of a bias lowering the measured tumour mean value. Of the three difference-image prior methods, DE-PML provides on average the best tradeoff between noise reduction in the background, noise reduction in the tumour and bias-mitigation in the tumour.

When considering the example reconstructed images at $\sigma_w = \infty$ (figure 11), the improvements are visually obvious. Compared to the MLEM reconstructions with the same level of counts, the difference-image priors produce visually superior images with reduced noise throughout the image. These noise levels are visually similar to the

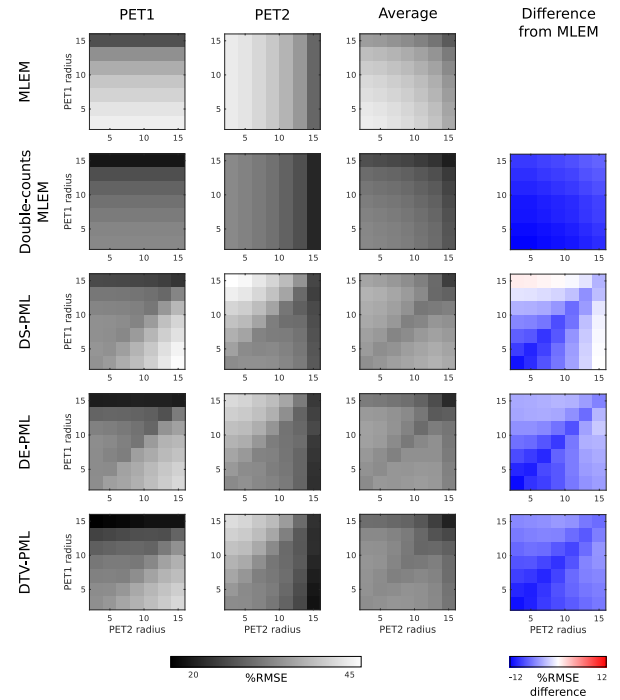


Figure 7. %RMSE values in PET1 and PET2 reconstructions in a region encompassing the tumour for varying tumour radii for: MLEM, double-counts MLEM, and DS-PML, DE-PML and DTV-PML with β values from table I. Tumour intensities were kept constant at 8 and 6 (relative to a nominal white matter uptake of 1). Also shown are average %RMSEs across PET1 and PET2 and the difference of these averages from the MLEM averages. Note that in all cases the analysis region was a dilation of the largest (15 mm) tumour, and so for smaller tumours the analysis region contained proportionally more background. It is apparent that across tumour sizes, the difference-image prior methods allow for good reconstruction of the region around the tumour, with error values generally falling between MLEM and double-counts MLEM reconstructions.

MLEM reconstructions with $5\times$ the data. In the tumour itself, the effects of the different methods manifest themselves. The DS-PML reconstruction provides a tumour that appears similar to the MLEM reconstruction with the same amount of data, due to the tendency of this method to leave regions of change relatively unaltered. For the DE-PML and DTV-PML methods, the tumour itself is visually improved compared to MLEM with the same number of counts, due to noise reduction in the tumour.

Finally, figure 12 shows the mean-squared-error (MSE, equal to the sum of voxel-wise bias squared and voxel-wise standard deviation squared) images for each of the methods. Compared to MLEM reconstructions with the same amount of data, the difference-image prior reconstructions reduce image-wide MSE, due to noise-reduction in shared regions. In the tumour the levels of error reduction vary between the proposed methods. The DS-PML method slightly reduces error compared to

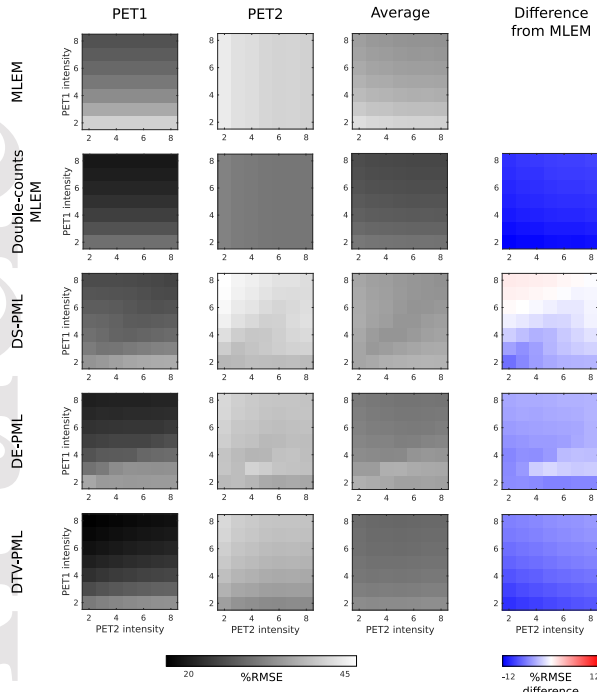


Figure 8. %RMSE values in PET1 and PET2 reconstructions in a region encompassing the tumour for varying tumour intensities for: MLEM, double-counts MLEM, and DS-PML, DE-PML and DTV-PML with β values as listed in table I. Tumour radii were kept constant at 15 mm and 8 mm. Also shown are average %RMSEs across PET1 and PET2 and the difference of these averages from the MLEM values. Results are similar to those seen in figure 7, with error levels generally between MLEM and double-counts MLEM, although both the DS-PML and DE-PML methods produce %RMSE values more similar to MLEM for a greater number of intensity changes.

the MLEM reconstructions, due to the small amount of noise-reduction that occurs when voxel-values are longitudinally biased offsetting the bias itself, as previously reported²⁶. For the DE-PML and DTV-PML methods, error reduction in the tumour is stronger. Here we see that despite the correct estimation of tumour means provided by the DE-PML method in figure 9, this method still produces higher voxel-wise errors in the tumour compared to DTV-PML. This suggests that the tumour means are maintained with DE-PML at the expense of higher levels of tumour voxel variance.

D. Experiment 4: Application to real data for head and neck cancer

The tumour mean and background CV for the real data experiment are shown in figure 13. Similarly to the results seen in the simulation experiments, increasing β reduces background noise while affecting the tumour mean. In the PET1 reconstructions, DS-PML and DTV-PML introduced a negative bias into the tumour

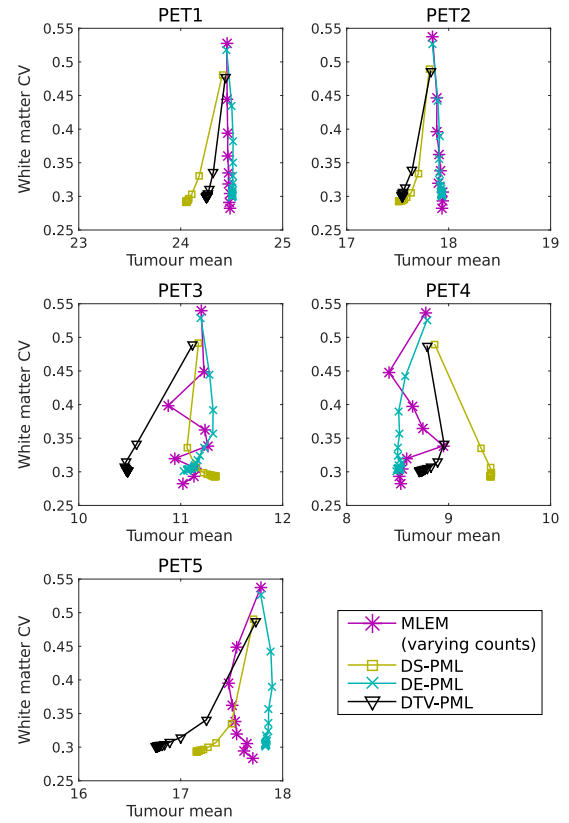


Figure 9. Tumour mean vs white matter CV for the five scans for DS-PML, DE-PML and DTV-PML as a function of σ_w . Penalty strengths, β , were set to the values shown in table I. Also shown is MLEM as a function of recorded counts, with the number of counts varied from $1\times$ to $5\times$ the number of counts used in the proposed methods. All the proposed methods are able to reduce the background noise considerably, at the cost of some level of bias in the reconstructions. This bias is smallest for the DE-PML method, which remains close to the MLEM reconstructions with higher levels of counts.

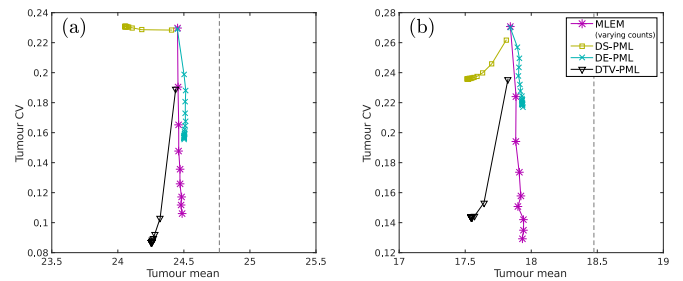


Figure 10. Tumour mean vs tumour CV for (a) PET1 and (b) PET2 reconstructions for the five-scan simulation experiment (PET3 to PET5 tumours have been omitted due to their small size prohibiting useful measures of tumour CV). For DS-PML, DE-PML and DTV-PML, β values are as listed in table I, and σ_w was varied between 0 and ∞ . For MLEM, the number of counts per dataset was varied from $1\times$ to $5\times$ the number used in the other reconstructions. For the DE-PML and DTV-PML methods, increasing σ_w reduced tumour noise while introducing some bias relative to MLEM reconstructions, whereas with DS-PML, particularly in PET1, the effect on tumour CV was weak.

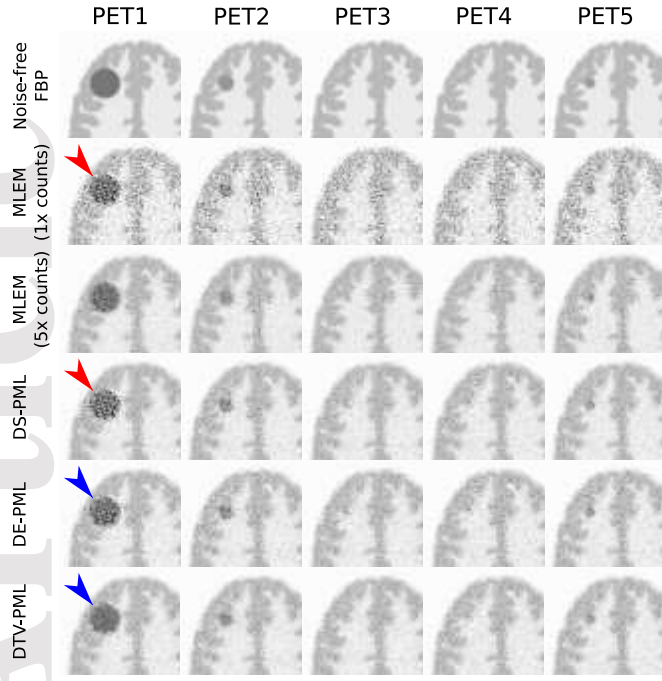


Figure 11. Example reconstructed images for the simulated five-scan reconstruction experiment. Each column shows the images from each of the scans. From top to bottom: noise-free FBP, MLEM with a counts level of 1, MLEM with $5\times$ the number of counts, DS-PML, DE-PML and DTV-PML. All of the difference-image prior reconstruction methods had $\sigma_w = \infty$ and β values are those listed in table I. The three difference-image priors reduce noise in the background of the image to levels equivalent to the $5\times$ counts MLEM, but differ in the appearance of the tumours, particularly in PET1, where DS-PML retains a similar tumour noise pattern to the MLEM reconstruction with the same number of counts (red arrows) and DE-PML and DTV-PML reduce the noise (blue arrows).

mean, whereas DE-PML produced a positive bias. In the PET2 reconstructions, DS-PML and DTV-PML produced tumour mean values similar to standard reconstruction techniques (i.e. with $\beta = 0$), which themselves were positively biased compared to the double-counts reference OSEM reconstruction. It was also observed that while the background CV in the PET1 images reduced to levels similar to the double-counts OSEM reconstruction, in agreement with the simulation study results, the PET2 background CV did not fall by the same amount, remaining considerably higher than the double-counts reconstruction even at the highest β values used.

Based on figure 13, β values were selected for the real data case. These values were 0.2 for DS-PML, 2000 for DE-PML, and 0.04 for DTV-PML. Figure 14 shows the reconstructed images for each of these methods, alongside the double-counts OSEM reconstruction and the OSEM reconstruction with the same data as the difference-image prior methods. It was noted that the tumour in the PET2 image consisted of two distinct peaks as opposed to the

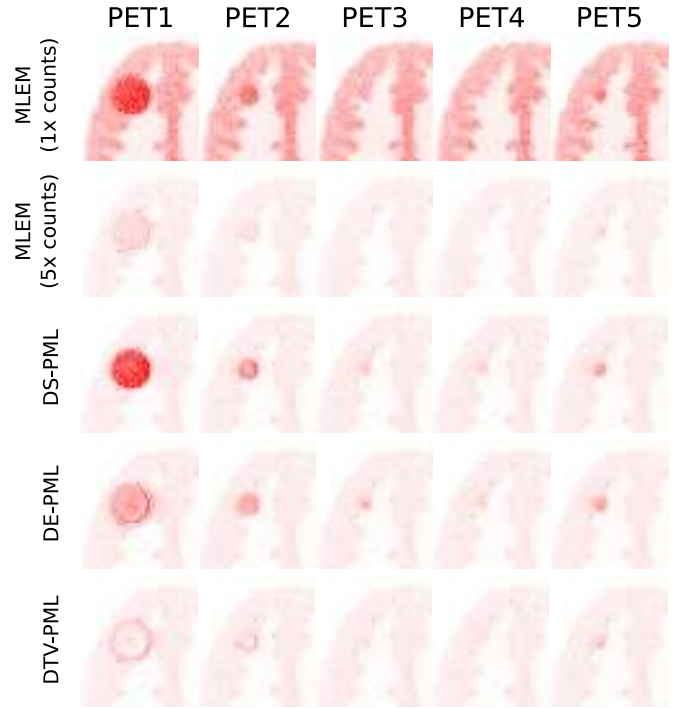


Figure 12. Mean-squared-error (MSE) maps around the tumour for MLEM with a counts level of 1, MLEM with $5\times$ the number of counts, and DS-PML, DE-PML and DTV-PML. β values for each of the difference-image priors are listed in table I. All of the difference-image prior reconstruction methods had $\sigma_w = \infty$. Note that MSE is calculated as the sum of bias squared plus standard deviation squared.

single peak observed in the PET1 images. The background noise reduction indicated by figure 13 is visible in the difference-image prior methods, particularly in the PET1 scans. The appearance of the tumours generally appears the same as the OSEM methods, although DTV-PML smooths the tumour slightly. The distinction between the three methods is most pronounced when considering the difference images though, where the mechanism of each of the three priors becomes apparent. The DS-PML method successfully sets a large portion of the difference image to zero (or close to zero), enhancing the visibility of the remaining changes. DE-PML has also set much of the difference image background to zero, with relatively few remaining intense voxels, mainly in the tumour. Lastly, the DTV-PML method has effectively smoothed the difference image. It is noted that in this real data case, imperfections in the alignment process have left other large changes present in the difference image, most noticeably around the cerebellum.

Finally, figure 15 shows the profiles through the tumour in the PET1, aligned PET2, and difference images for each reconstruction method. In the PET1 DE-PML reconstruction (figure 15(a)), the tumour peak is considerably higher than the other reconstruction methods, reflecting the positive bias observed in figure 13. In the PET2 reconstructions (figure 15(b)), the overestimation

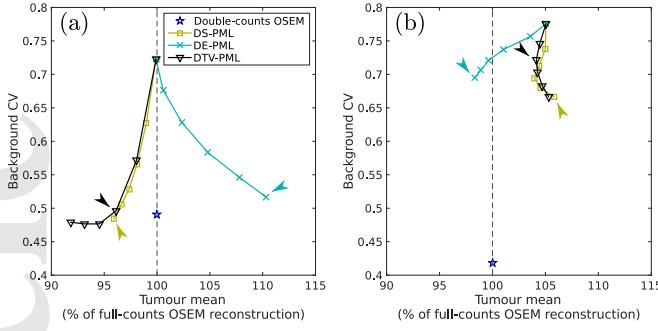


Figure 13. Tumour mean vs background CV as a function of β for DS-PML (yellow squares), DE-PML (cyan crosses) and DTV-PML (black triangles) for the real data experiment. Tumour mean values are shown as a percentage of the double-counts OSEM reconstruction (blue stars). In the PET1 reconstructions (a), when using the difference-image prior methods the CV in the background is reduced to levels similar to the double-counts reconstruction, however in PET2, the background CV does not reduce so greatly. Arrows for each method indicate the β values used in the reconstructions shown in figures 14 and 15.

of the tumour mean by OSEM, DS-PML and DTV-PML is shown in the left peak of the tumour profile. The difference image profiles (figure 15(c)) show the overestimation of the difference provided by DE-PML, and also the effect of each method in the observed differences in the smaller peak of the PET2 tumour, which was more intense in PET2 than in PET1 (red circle). DS-PML and DE-PML suppress this small real change relative to the OSEM reconstructions, whereas DTV-PML better preserves it.

IV. DISCUSSION

In this work we have built upon the concept of simultaneous reconstruction of longitudinal PET datasets via the use of difference-image penalty terms. Specifically, we have proposed novel penalties that encourage low entropy difference images (DE-PML) and difference images with sparse spatial gradients (DTV-PML). The results of 2D simulation study experiments have shown that using longitudinal difference-image priors to reconstruct S longitudinal scans can reduce image noise up to levels typically achieved by standard reconstructions using $S \times$ the number of recorded counts. However, there is a tradeoff between the level of background noise reduction achieved and the reconstruction of regions of change between scans.

In particular, the results from Experiment 1 showed that all of the proposed penalty terms introduce some level of bias in terms of estimation of tumour mean values with sufficiently high penalty strengths (figure 1). Nonetheless, the entropy-based method DE-PML was observed to produce the lowest levels of bias when using a double-counts MLEM as the reference. On the

other hand, when considering reconstruction error in and around the tumour, the DTV-PML method produced the lowest %RMSE values, beating even the DE-PML method (figure 2). Given that the DE-PML method reconstructs tumour means more accurately, the improvement in %RMSE seen with the DTV-PML method results from noise reduction within the tumour, an observation reinforced by the lower voxel-wise variance observed in figure 6. Overall though, the level of bias introduced by any of the proposed methods was small, and inspection of the longitudinal trends of the tumour mean (figure 5) shows that for this particular case (intensity reduction of 25% and radius reduction of 47%), even a relatively strong regularisation in terms of high levels of background noise reduction allows for clear discernment of the behaviour of the tumour.

When the regularisation level was held constant and tumour response varied, the performance of the proposed methods was largely stable. At large radius changes (with fixed intensities), error levels began to rise, with DTV-PML producing the lowest errors of the three penalties (figure 7). This is due to the noise reduction properties observed in Experiment 1. However, in general, with any of the three methods error levels in the tumour as a function of response are more complicated than for MLEM reconstructions. Before use in a clinical setting, the proposed methods would need to be validated across a representative, application-specific set of tumour responses to ensure that they are robust to the range of changes anticipated in that context. Such a validation study is beyond the scope of this present work which focuses primarily on the proposal of new reconstruction methodologies.

Extending the difference-image prior methods to the five-scan 2D case reinforced the results seen previously with the two-scan case. In particular, the tradeoff between background noise reduction and tumour bias was shown to hold true for varying σ_w , which controlled the coupling between scans throughout the longitudinal series (figure 9). At $\sigma_w = \infty$, maximum noise reduction was achieved at the expense of bias levels similar to those observed in the two-scan case. Also reflected from the two-scan case was the ability of the DE-PML method to more accurately reconstruct tumour means, and the error reduction of DTV-PML due to noise reduction in the tumour (figure 12). In the five-scan case, the visual impact of using difference-image priors is stronger in the reconstructed images due to the greater amount of data that is available to be shared between the scans.

Overall, the results of the real data case study reflected those of the simulation experiments, with increasing β reducing background image noise and preserving tumour appearance. Again, some bias was introduced into the tumour mean relative to the standard OSEM reconstructions. However, the real data experiment also highlighted some of the issues with the three difference-image priors, particularly when there are additional changes between the images beyond those caused by the tumour response,

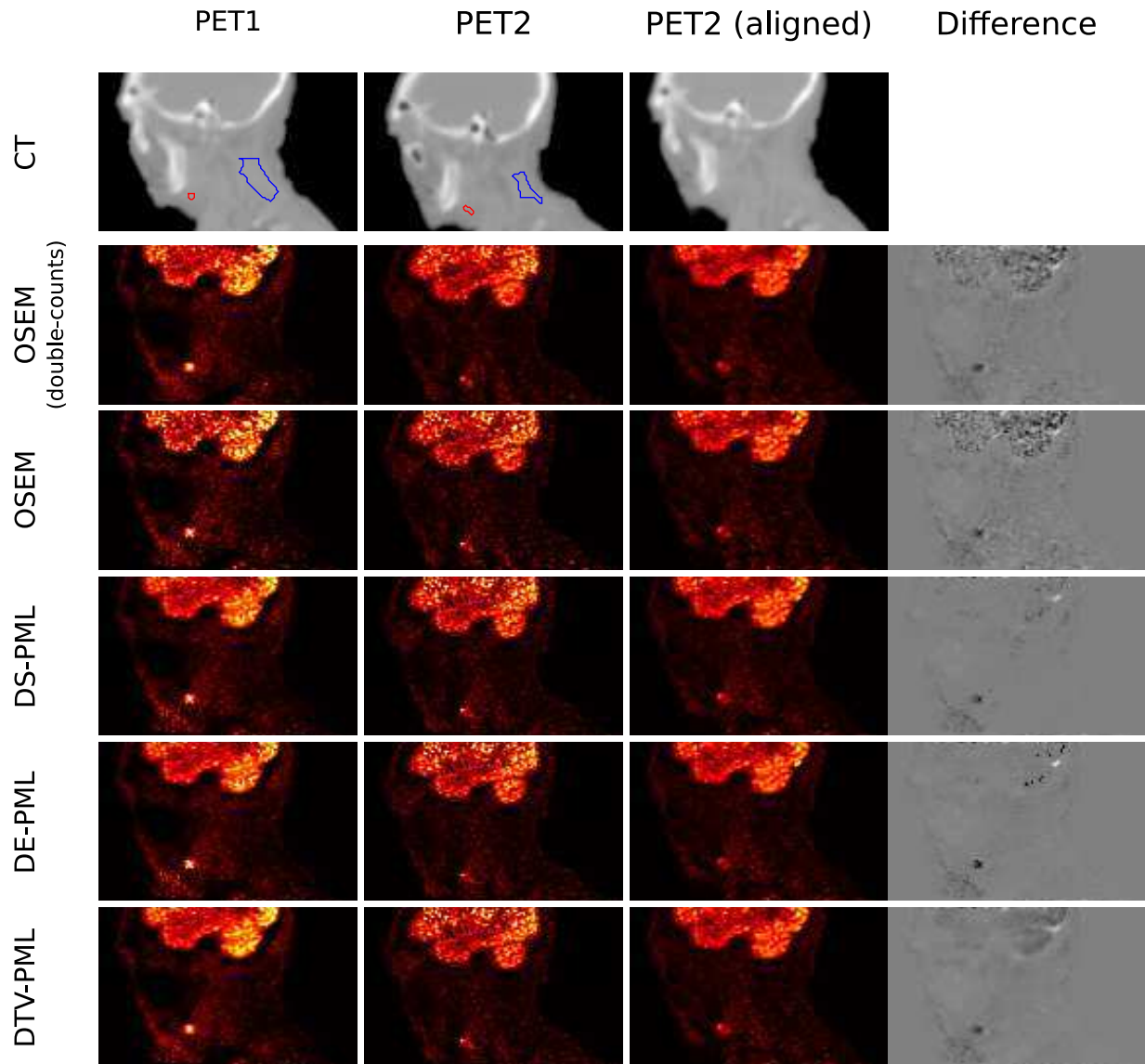


Figure 14. Example images for the real data case study. From left to right: PET1 images in their native space; PET2 images in their native space; PET2 images aligned to the PET1 image space; and difference images in the PET1 image space (i.e. column 3 minus column 1). Top row shows CT images with tumour (red) and background (blue) regions of interest shown for reference. PET reconstructions include the double-counts OSEM reconstructions, OSEM reconstructions, and reconstructions using DS-PML, DE-PML and DTV-PML, each with β values as indicated in figure 13. These latter three difference-image prior methods reduce noise in the background of the image while maintaining the visual appearance of the tumour. Each also displays a distinctive difference image reflecting the characteristics encouraged by each prior.

such as those that were observed in the brain of the patient. For the DS-PML method, the existence of these additional strong changes could be the cause of the suppression of the tumour differences caused by the appearance of a second tumour peak in the PET2 scan (figure 15, red circle). Because the amplitude of this valid tumour change was lower than the differences observed elsewhere in the image, it was suppressed in place of these other changes in order to encourage the sparsity sought by the prior term. Likewise, the existence of the changes within the brain is suspected to be the cause of the larger biases observed with the DE-PML method; by attempt-

ing to produce a low entropy difference image, similar changes can be encouraged to have the same value, even over large spatial distances. This is reflected in figure 14, where it can be seen that the changes in the brain and the tumour in the DE-PML difference image seem to have similar values, despite them appearing distinct in the other difference images. This issue could be alleviated in future by adopting a local entropy prior on the difference image, encouraging spatially distant parts of the image to possess their own low-entropy distributions without allowing them to interfere with each other. Contrary to the other two methods, the DTV-PML method

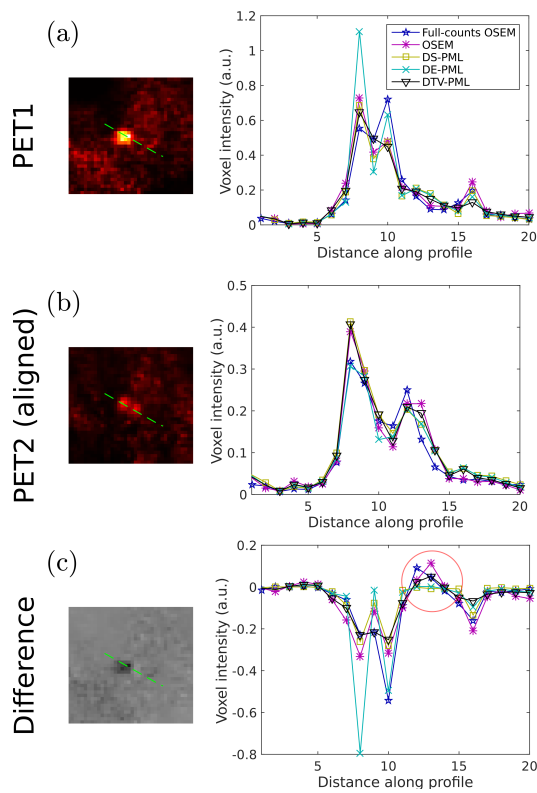


Figure 15. Line profiles through the tumour for (a) the PET1 images, (b) the aligned PET2 images, and (c) the difference images (bottom). The red circle highlights differences due to a region that increased in uptake from PET1 to PET2 that was suppressed by the DS-PML and DE-PML methods but retained by the DTV-PML method.

seems to deal better with the existence of other changes in the difference image, maintaining the second tumour peak that was suppressed by the other two methods (figure 15, red circles). However, the DTV-PML method still introduced bias into the tumour reconstructions because of spatial smoothing applied by the penalty in the difference image domain. This reflects the results seen in the simulation studies (e.g. figure 3(a)), where the tumour was smoothed relative to the standard MLEM reconstructions.

The proposal of longitudinal image reconstruction for PET images is, as far as the authors are aware, still a novel concept previously proposed only by our group²⁶. In MR imaging however, the coupling of longitudinal datasets has been previously proposed, via the use of adaptive data acquisition and penalisation of differences between the follow-up image and the baseline image in the reconstruction^{32,33}. While there are similarities between this method and the ones we describe here and previously²⁶, coupling of MR datasets is distinct from the PET case since in MR imaging information loss in the form of data undersampling results in image artefacts and not necessarily increased image noise. Therefore, in MR, it is possible to have a single high quality

baseline scan to be able to improve the follow-up scan in a manner analogous to guided reconstructions in PET. On the other hand, in PET imaging, the noise levels are almost always high. In this case, reconstructing the follow-up scan by penalising differences with respect to a pre-reconstructed baseline image would not be useful. By penalising both images simultaneously as in our proposed framework, noise can be effectively suppressed at early iterations, achieving the inter-scan transfer of information for the PET case.

Another area of research related to longitudinal reconstruction is that of dynamic, or 4D PET image reconstruction. In 4D reconstruction, the temporal behaviour of a tracer distribution is required, and there exist many methodologies for estimating this³⁴. However, there are notable differences between 4D reconstruction methods and longitudinal reconstruction methods. Firstly, 4D reconstructions generally have access to all data from within a dynamic acquisition, which ensures that timeframes are adjacent. This means that it is easier to encourage temporal similarity between frames, since it is safe to assume that no large changes have occurred. In the longitudinal case, the scans could quite feasibly be separated by months, which is ample time for large changes to occur. Secondly, because the overall aim of 4D reconstruction is often to estimate kinetic parameters according to pharmacokinetic compartmental models, there can be a greater level of robustness to biases in individual isolated voxel values that can be caused by the model fitting. On the other hand however, in longitudinal studies the exact values of voxels and ROIs can be very important for a given single scan, often more so than the case of a single frame within a 4D reconstruction. Finally, in 4D PET contexts, the number of time frames (commonly pseudo-continuous temporal sampling is possible when reconstructing in list-mode; or at least > 10 timeframes are used) is considerably higher than the typical number of scans available for longitudinal studies (2 is standard; 5 would be considered high). This limitation of longitudinal studies means that it is more difficult to devise meaningful longitudinal models which would be suitable for direct incorporation into the image reconstruction method.

The methods and results presented in this work have, of course, their limitations. Firstly, and most importantly, the introduction of bias into the tumour region could be a cause for concern. However, as shown in this work, these biases are relatively small compared to typically expected variations due to tumour progression or response. Furthermore, the amount of bias that is acceptable will be application specific. For example, in tumour detection contexts, the benefit of background noise reduction provided by our proposed methods might be considered worth the bias introduced into the tumour. On the other hand, in tumour response scans where the threshold for responder or non-responder is very fine these methods might be of limited benefit (or at least lower penalty strengths would need to be used).

As highlighted by the real data experiment, another limitation of the presented difference-image prior reconstruction methods is the need for high-quality alignment fields to be used during the reconstruction. Errors in alignment fields introduce additional features into the difference images, which can adversely impact the effectiveness of the priors as described above. In this work we used non-rigid deformation fields estimated using a demons-based method, which resulted in difference images that contained some residual changes that were not due to the tumour. Future work would require refinement of the registration process by considering other registration algorithms and optimising all relevant parameters, although the most appropriate registration algorithm will depend on the application. In this respect, previous work aiming to provide regional or voxel-level analysis of tumour differences^{35–38} might be informative as to the applications that are most well-developed for incorporation of the methods presented in this paper. Another opportunity for improvement, with the increased usage of simultaneous PET-MR scanners, is the use of MR images, with superior soft tissue contrast over CT, to perform the registration. Furthermore, incorporating the estimation of motion fields during PET image reconstruction is a growing area of research^{39–41}, which could have potential for refining alignment fields estimated from other imaging modalities.

Ultimately, however, the methods presented in this work need to be tested in an application-specific manner to ascertain which areas could benefit most from their use. This would include testing over many representative cases to allow parameter values like β and σ_w (or rules for setting them) to be derived based on a population, rather than the limited single object used in the simulations in this work. It is important to note that while this work has focused on oncology treatment response scans as an application of longitudinal imaging, there are many others where difference-image priors could be of use, such as in the long-term observation of neuro-degenerative disorders^{42,43}, ictal and interictal subtraction imaging in epilepsy^{44,45}, and functional brain studies that explore changes in brain activation arising from drugs or other stimuli^{46–48}.

V. CONCLUSIONS

Longitudinal PET scans are often acquired in order to observe and quantify physiological and molecular changes *in vivo* over extended time scales. Typically these scans are reconstructed independently using standard PET reconstruction algorithms such as MLEM. Simultaneous reconstruction of longitudinal PET datasets is a newly proposed area of research which has opened up many possibilities for improvement of longitudinal PET images. This work builds on the promising previous preliminary results by proposing additional difference-image priors and evaluating their performance. In simulation studies,

results showed that considerable noise reduction can be achieved across background portions of the image. Performance of the reconstruction methods within the tumour itself vary, with encouragement of a low-entropy difference image maintaining low levels of bias in the tumour, while encouraging difference images with few spatial gradients provides the lowest reconstruction errors by way of noise reduction at the cost of induced tumour bias. Application to a real data case study showed similar improvements, but highlighted the need for high-quality alignment fields to allow the proposed methods to perform optimally.

Future work will require exploration of the issue of mis-registration of longitudinal scans by exploring other registration algorithms and optimising parameters. Furthermore, evaluation of the performance of the proposed methods in an application-specific manner will need to be carried out in order to ascertain which areas of longitudinal PET imaging, not limited to oncology, will benefit from simultaneous image reconstruction. Ultimately, the coupling of longitudinal datasets during image reconstruction has the potential to improve longitudinal PET images ‘for free’, if current injected doses are used, or to reduce injected dose (and therefore the patient’s absorbed dose) by a factor approximately equal to the number of longitudinal scans and maintain current levels of image quality.

ACKNOWLEDGEMENTS

This work was funded by the King’s College London and Imperial College London EPSRC Centre for Doctoral Training in Medical Imaging [EP/L015226/1] and supported by both the EPSRC [EP/M020142/1] and the Wellcome EPSRC Centre for Medical Engineering at King’s College London [WT 203148/Z/16/Z]. The authors would also like to thank Zacharias Chalamalakos for providing the patient data and Abolfazl Mehranian for providing code for reconstructing Siemens Biograph mCT data. According to the EPSRC’s policy framework on research data, figures and data supporting this study are openly available at <https://doi.org/10.5281/zenodo.1219686>.

DISCLOSURE OF CONFLICTS OF INTEREST

The authors have no relevant conflicts of interest to disclose.

REFERENCES

- ¹Zhu A, Lee D, Shim H. Metabolic Positron Emission Tomography Imaging in Cancer Detection and Therapy Response *Semin. Oncol.*. 2011;38:55–69.

- ²Kishino T, Hoshikawa H, Nishiyama Y, Yamamoto Y, Mori N. Usefulness of 3'-Deoxy-3'-¹⁸F-Fluorothymidine PET for Predicting Early Response to Chemoradiotherapy in Head and Neck Cancer *J. Nucl. Med.*. 2012;1-7.
- ³Lordick F, Ott K, Krause B-J, et al. PET to assess early metabolic response and to guide treatment of adenocarcinoma of the oesophagogastric junction: the MUNICON phase II trial *Lancet Oncol.*. 2007;8:797-805.
- ⁴Georg P, Andrzejewski P, Baltzer P, et al. Changes in Tumor Biology During Chemoradiation of Cervix Cancer Assessed by Multiparametric MRI and Hypoxia PET *Mol. Imaging Biol.*. 2018;20:160-169.
- ⁵Stroobants S, Goeminne J, Seegers M, et al. 18FDG-Positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec®) *Eur. J. Cancer.* 2003;39:2012-2020.
- ⁶Ishimori T, Saga T, Nagata Y, et al. ¹⁸F-FDG and ¹¹C-methionine PET for evaluation of treatment response of lung cancer after stereotactic radiotherapy *Ann. Nucl. Med.*. 2004;18:669-674.
- ⁷Dercle L, Seban R-D, Lazarovici J, et al. ¹⁸F-FDG PET and CT Scans Detect New Imaging Patterns of Response and Progression in Patients with Hodgkin Lymphoma Treated by Anti-Programmed Death 1 Immune Checkpoint Inhibitor *J. Nucl. Med.*. 2018;59:15-24.
- ⁸Cheebsumon P, Velasquez LM, Hoekstra CJ, et al. Measuring response to therapy using FDG PET: semi-quantitative and full kinetic analysis *Eur. J. Nucl. Med. Mol. Imaging.* 2011;38:832-842.
- ⁹Brun E, Kjellén E, Tennvall J, et al. FDG PET studies during treatment: Prediction of therapy outcome in head and neck squamous cell carcinoma *Head Neck.* 2002;24:127-135.
- ¹⁰Shepp LA, Vardi Y. Maximum Likelihood Reconstruction for Emission Tomography *IEEE Trans. Med. Imag.*. 1982;1:113-122.
- ¹¹Hudson HM, Larkin RS. Accelerated Image Reconstruction Using Ordered Subsets of Projection Data *IEEE Trans. Med. Imag.*. 1994;13:601-609.
- ¹²Qi J, Leahy RM. Iterative reconstruction techniques in emission computed tomography *Phys. Med. Biol.*. 2006;51:R541.
- ¹³Wang G, Qi J. Penalized Likelihood PET Image Reconstruction Using Patch-Based Edge-Preserving Regularization *IEEE Trans. Med. Imag.*. 2012;31:2194-2204.
- ¹⁴Wang G, Qi J. Edge-Preserving PET Image Reconstruction Using Trust Optimization Transfer *IEEE Trans. Med. Imag.*. 2015;34:930-939.
- ¹⁵Teng Y, Sun H, Guo C, Kang Y. ADMM-EM Method for L1-Norm Regularized Weighted Least Squares PET Reconstruction *Comput. Math. Methods Med.*. 2016;2016.
- ¹⁶Somayajula S, Panagiotou C, Rangarajan A, Li Q, Arridge SR, Leahy RM. PET Image Reconstruction Using Information Theoretic Anatomical Priors *IEEE Trans. Med. Imag.*. 2011;30:537-549.
- ¹⁷Vunckx K, Atré A, Baete K, et al. Evaluation of Three MRI-Based Anatomical Priors for Quantitative PET Brain Imaging *IEEE Trans. Med. Imag.*. 2012;31:599-612.
- ¹⁸Bai B, Li Q, Leahy RM. MR-Guided PET Image Reconstruction *Semin. Nucl. Med.*. 2013;43:30-44.
- ¹⁹Nguyen VG, Lee SJ. Incorporating Anatomical Side Information Into PET Reconstruction Using Nonlocal Regularization *IEEE Trans. Med. Imag.*. 2013;22:3961-3973.
- ²⁰Ehrhardt MJ, Markiewicz P, Liljeroth M, et al. PET Reconstruction With an Anatomical MRI Prior Using Parallel Level Sets *IEEE Trans. Med. Imag.*. 2016;35:2189-2199.
- ²¹Novosad P, Reader AJ. MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions *Phys. Med. Biol.*. 2016;61:4624-4645.
- ²²Mehranian A, Belzunce MA, Niccolini F, et al. PET image reconstruction using multi-parametric anato-functional priors *Phys. Med. Biol.*. 2017;62:5975.
- ²³Ehrhardt MJ, Thielemans K, Pizarro L, et al. Joint reconstruction of PET-MRI by exploiting structural similarity *Inverse Probl.*. 2015;31:015001.
- ²⁴Mehranian A, Belzunce MA, Prieto C, Hammers A, Reader AJ. Synergistic PET and SENSE MR image reconstruction using joint sparsity regularization [published online ahead of print April 18, 2017]. *IEEE Trans. Med. Imag.*. doi: 10.1109/TMI.2017.2691044.
- ²⁵Knoll F, Holler M, Koesters T, Otazo R, Bredies K, Sodickson DK. Joint MR-PET Reconstruction Using a Multi-Channel Image Regularizer *IEEE Trans. Med. Imag.*. 2017;36:1-16.
- ²⁶Ellis S, Reader AJ. Simultaneous maximum *a posteriori* longitudinal PET image reconstruction *Phys. Med. Biol.*. 2017;62:6963.
- ²⁷Green PJ. On Use of the EM for Penalized Likelihood Estimation *J. R. Stat. Soc. Ser. B Stat. Methodol.*. 1990;52:443-452.
- ²⁸Panin VY, Zeng GL, Gullberg GT. Total Variation Regulated EM Algorithm *IEEE Trans. Nucl. Sci.*. 1999;46:2202-2210.
- ²⁹Rahmim A, Dinelle K, Cheng JC, et al. Accurate Event-Driven Motion Compensation in High-Resolution PET Incorporating Scattered and Random Events *IEEE Trans. Med. Imag.*. 2008;27:1018-1033.
- ³⁰Burger C, Goerres G, Schoenes S, Buck A, Lonn A, Schulthess G. PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients *Eur. J. Nucl. Med. Mol. Imaging.* 2002;29:922-927.
- ³¹Mehranian A, Kotasidis F, Zaidi H. Accelerated time-of-flight (TOF) PET image reconstruction using TOF bin subsetization and TOF weighting matrix pre-computation *Phys. Med. Biol.*. 2016;61:1309-1331.
- ³²Weizman L, Eldar YC, Ben Bashat D. Compressed sensing for longitudinal MRI: An adaptive-weighted approach *Med. Phys.*. 2015;42:5195-5208.
- ³³Kang R, Qu G, Cao B, Yan L. Combined Similarity to Reference Image with Joint Sparsifying Transform for Longitudinal Compressive Sensing MRI *Math. Probl. Eng.*. 2016;2016:1-12.
- ³⁴Reader AJ, Verhaeghe J. 4D image reconstruction for emission tomography *Phys. Med. Biol.*. 2014;59:R371-R418.
- ³⁵Necib H, Garcia C, Wagner A, et al. Detection and Characterization of Tumor Changes in 18F-FDG PET Patient Monitoring Using Parametric Imaging *J. Nucl. Med.*. 2011;52:354-361.
- ³⁶Schreibmann E, Waller AF, Crocker I, Curran W, Fox T. Voxel clustering for quantifying PET-based treatment response assessment *Med. Phys.*. 2013;40:012401-01-012401-12.
- ³⁷Vera P, Dubray B, Palie O, et al. Monitoring tumour response during chemo-radiotherapy: a parametric method using FDG-PET/CT images in patients with oesophageal cancer *EJNMMI Res.*. 2014;4:1-12.
- ³⁸Kerner GSMA, Fischer A, Koole MJB, Pruim J, Groen HJM. Evaluation of elastix-based propagated align algorithm for VOI- and voxel-based analysis of longitudinal 18F-FDG PET/CT data from patients with non-small cell lung cancer (NSCLC) *EJNMMI Res.*. 2015;5:1-15.
- ³⁹Jacobson MW, Fessler JA. Joint estimation of image and deformation parameters in motion-corrected PET in 2003 *IEEE Nuclear Science Symposium. Conference Record*;5:3290-3294 2003.
- ⁴⁰Blume M, Martinez-Moller A, Keil A, Navab N, Rafecas M. Joint Reconstruction of Image and Motion in Gated Positron Emission Tomography *IEEE Trans. Med. Imag.*. 2010;29:1892-1906.
- ⁴¹Kalantari F, Li T, Jin M, Wang J. Respiratory motion correction in 4D-PET by simultaneous motion estimation and image reconstruction (SMEIR) *Phys. Med. Biol.*. 2016;61:5639.
- ⁴²Mosconi L, Mistur R, Switalski R, et al. FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease *Eur. J. Nucl. Med. Mol. Imaging.* 2009;36:811-822.
- ⁴³Jagust WJ, Bandy D, Chen K, et al. The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core *Alzheimers. Dement.*. 2010;6:221-229.
- ⁴⁴Desai A, Bekelis K, Thadani VM, et al. Interictal PET and ictal subtraction SPECT: Sensitivity in the detection of seizure

foci in patients with medically intractable epilepsy *Epilepsia*. 2013;54:341–350.

⁴⁵Perissinotti A, Setoain X, Aparicio J, et al. Clinical Role of Subtraction Ictal SPECT Coregistered to MR Imaging and 18F-FDG PET in Pediatric Epilepsy *J. Nucl. Med.*. 2014;55:1099–1105.

⁴⁶Corbetta M, Miezin FM, Shulman GL, Petersen SE. A PET Study of Visuospatial Attention *J. Neurosci.*. 1993;13:1202–1226.

⁴⁷Schlindwein P, Buchholz H-G, Schreckenberger M, Bartenstein P, Dieterich M, Birklein F. Sympathetic activity at rest and mo-

tor brain areas: FDG-PET study *Auton. Neurosci. Basic Clin.*. 2008;143:27–32.

⁴⁸Salimpoor VN, Benovoy M, Larcher K, Alain D, Zatorre RJ. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music *Nat. Neurosci.*. 2011;14:257–262.